

University of Groningen

Advances in artificial neural networks, machine learning and computational intelligence

Aioli, Fabio; Biehl, Michael; Oneto, Luca

Published in:
Neurocomputing

DOI:
[10.1016/j.neucom.2018.01.090](https://doi.org/10.1016/j.neucom.2018.01.090)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Aioli, F., Biehl, M., & Oneto, L. (2018). Advances in artificial neural networks, machine learning and computational intelligence. *Neurocomputing*, 298, 1-3. <https://doi.org/10.1016/j.neucom.2018.01.090>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Advances in artificial neural networks, machine learning and computational intelligence

Selected papers from the 25-th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017)

This special issue of Neurocomputing presents 13 original articles that are extended versions of selected papers from the 25-th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017), a major event for researchers in the field of artificial neural networks and related topics. This single track conference is held annually in Bruges, Belgium, a UNESCO World Heritage Site by one of the most beautiful medieval city centers in Europe. It is organized in collaboration by UCL (Université Catholique de Louvain – Louvain-la-Neuve) and KULeuven (Katholiek Universiteit – Leuven) and is steered by Prof. Michel Verleysen from UCL. In addition to regular sessions, the conference regularly welcomes special sessions organized by renowned scientists in their respective fields. These sessions focus on particular topics, for instance deep learning, kernel methods, randomized learning approaches, biomedical and environmental data analysis, clustering, data visualization, and big data analytics.

The contributions in this special issue show that ESANN covers a broad range of topics in neural computing and neuroscience, from theoretical aspects to state-of-the-art applications. More than 100 researchers from 20 countries participated in the 25-th ESANN in April 2017. Around 100 oral and poster communications have been presented this year. Based on the reviewers' and special session organizers' recommendations, as well as on the quality of the oral presentations at the conference, a number of authors were invited to submit an extended version of their conference paper for this special issue of Neurocomputing. All extended manuscripts were thoroughly reviewed once more by at least two independent experts and the 13 articles presented in this volume were accepted for publication. They can be grouped as follows.

1) Randomized Learning

i) Peter Tiño

Asymptotic Fisher Memory of Randomized Linear Symmetric Echo State Networks

This paper studies the asymptotic properties of Fisher memory of linear Echo State Networks with randomized symmetric state space coupling. In particular, two reservoir constructions are considered: a more direct dynamic coupling construction using a class of Wigner matrices, and a positive semi-definite dynamic coupling obtained

as a product of unconstrained stochastic matrices. Paper shows that the maximal Fisher memory is achieved when the input-to-state coupling is collinear with the dominant eigenvector of the reservoir coupling matrix. In the case of Wigner reservoirs paper shows that as the system size grows, the contribution to the Fisher memory of self-coupling of reservoir units is negligible. The paper also proves that when the input-to-state coupling is collinear with the sum of eigenvectors of the state space coupling, the expected normalized memory is four and eight time smaller than the maximal memory value for the Wigner and product constructions, respectively.

- ii) Davide Bacciu, Michele Colombo, Davide Morelli, and David Plans
Randomized Neural Networks for Preference Learning with Physiological Data

The paper discusses the use of randomized neural networks to learn a complete ordering between samples of heart-rate variability data by relying solely on partial and subject-dependent information concerning pairwise relations between samples. Paper compares two approaches, i.e. Extreme Learning Machines and Echo State Networks and additionally, it introduces a weight sharing architecture and a preference learning error function whose performance is compared with a standard architecture realizing pairwise ranking as a binary-classification task. The models are evaluated on real-world data from a mobile application realizing a guided breathing exercise, using a dataset of over 54 thousand exercising sessions. Results show how a randomized neural model processing information in its raw sequential form can outperform its vectorial counterpart, increasing accuracy in predicting the correct sample ordering by about 20%. Furthermore, the experiments highlight the importance of using weight sharing architectures to learn smooth and generalizable complete orders induced by the preference relation.

- iii) Randomized Learning: Generalization Performance of Old and New Theoretically Grounded Algorithms

Luca Oneto, Francesca Cipollini, Sandro Ridella, and Davide Anguita

This paper develops tighter Differential Privacy based randomized model generalization bounds, which improve over the current state-of-the-art ones, based on the PAC-Bayes and Differential Privacy theories, both in terms of constants and rate of convergence. Moreover, the paper also proves that some old and new randomized algorithms show better generalization performances with respect to their non-private counterpart, if the Differential Privacy is exploited for assessing their generalization ability. Results on a series of algorithms and real world problems show the practical validity of the

achieved theoretical results.

2) Neural Networks

- i) Claudio Gallicchio, Alessio Micheli, and Luca Silvestri

Local Lyapunov Exponents of Deep Echo State Networks

This paper investigates the deep Echo State Networks model under a dynamical system perspective, aiming at characterizing the important aspect of stability of layered recurrent dynamics excited by external input signals. For this purpose, authors develop a framework based on the study of the local Lyapunov exponents of stacked recurrent models, enabling the analysis and control of the resulting dynamical regimes. Results show that when recurrent units are organized in layers, then the resulting network intrinsically develops a richer dynamical behavior that is naturally driven closer to the edge of criticality. This characterization makes the layered design more effective with respect to the shallow counterpart with the same number of units, as confirmed by the experiments on the short-term Memory Capacity task.

- ii) Siamak Mehrkanoon and Johan A. K. Suykens

Deep Hybrid Neural-Kernel Networks using Random Fourier Features

This contribution introduces a novel hybrid deep neural kernel framework. The proposed deep learning model combines a neural network based architecture with a kernel based model. The paper exploits an explicit feature map based on random Fourier features, in order to make the transition between the two architectures more straightforward as well as making the model scalable to large datasets by solving the optimization problem in the primal. The proposed framework is considered as the first building block for the development of even deeper models and more advanced architectures. Results show an improvement over shallow models and the standard non-hybrid neural networks architecture on several medium to large scale real-life datasets.

3) Regression and Classification

- i) Henry Reeve and Gavin Brown

Diversity and Degrees of Freedom in Regression Ensembles

This paper establishes a connection between diversity and degrees of freedom in ensemble methods, showing that diversity may be viewed as a form of inverse regularisation. This is achieved by focusing on a previously published algorithm Negative Correlation Learning, in which model diversity is explicitly encouraged through a diversity

penalty term in the loss function. The authors derive an exact expression for the effective degrees of freedom in a Negative Correlation Learning ensemble with fixed basis functions, showing that it is a continuous, convex and monotonically increasing function of the diversity parameter. This work demonstrates a connection to Tikhonov regularisation and shows that, with an appropriately chosen diversity parameter, a Negative Correlation Learning ensemble can always outperform the unregularised ensemble in the presence of noise.

- ii) Christina Göpfert, Lukas Pfannschmidt, Jan P. Göpfert, and Barbara Hammer

Interpretation of Linear Classifiers by Means of Feature Relevance Bounds

In this paper, the authors attempt to unify commonly used concepts of feature relevance, an often discussed but frequently poorly defined subject, and feature selection to give an overview of the main questions and results in finding all the relevant features for a classification problem. The authors formalize two interpretations of the all-relevant problem and propose a polynomial method to approximate one of them for the important hypothesis class of linear classifiers, which also enables a distinction between strongly and weakly relevant features.

4) Learning From Graphs

- i) Dhanesh Ramachandram, Michal Lisicki, Timothy J Shields, Mohamed R. Amer, and Graham W. Taylor

Bayesian Optimization on Graph-structured Search Spaces: Optimizing Deep Multimodal Fusion Architectures

A popular test bed for deep learning has been multimodal recognition of human activity or gesture involving diverse inputs such as video, audio, skeletal pose and depth images. Deep learning architectures have excelled on such problems due to their ability to combine modality representations at different levels of nonlinear feature extraction. However, designing an optimal architecture in which to fuse such learned representations has largely been a non-trivial human engineering effort. The authors treat fusion structure optimization as a hyperparameter search and cast it as a discrete optimization problem under the Bayesian optimization framework. They propose two methods to compute structural similarities in the search space of tree-structured multimodal architectures and demonstrate their effectiveness on two challenging multimodal human activity recognition problems.

- ii) Fabrizio Costa, Dinh Tran Van, and Alessandro Sperduti

The Conjunctive Disjunctive Graph Node Kernel for Disease Gene Prioritization

A high number of disease gene prioritization methods have been proposed since they play an important role in disclosing the relation between genes and disease. Among them, graph-based methods are the most promising paradigms due to their ability to naturally represent many types of relations using a graph representation. One key factor of success of graph-based learning methods is the definition of a proper graph node similarity measure, normally measured by graph node kernels. However, most approaches share two common limitations: first, they are based on the diffusion phenomenon which does not effectively exploit the nodes' context; second, they are not able to process the auxiliary information associated to graph nodes. In this paper, the authors propose an efficient graph node kernel, based on graph decompositions, that not only is able to effectively take into account the nodes' context, but also to exploit additional information available on graph nodes. An empirical evaluation in terms of several biological databases shows that the proposal of the authors achieves state-of-the-art results.

5) **Applications**

- i) Raul Barbosa, Douglas O. Cardoso, Diego Carvalho, and Felipe M França

Weightless Neuro-Symbolic GPS Trajectory Classification

This paper presents a framework for dealing with the problem of GPS trajectory classification in the context of the Rio de Janeiro's public transit system (with hundreds or more classes). Such framework combines the versatile WiSARD classifier with a set of rules defined a priori, resulting in a neuro-symbolic learning system with very interesting characteristics and cutting-edge performance. The authors have also verified the influence of different binarization methods in order to adapt raw data to WiSARD, which feeds from binary data only. The ideas presented in the paper were tested against a large data set of trajectories of buses from the city of Rio de Janeiro.

- ii) Emeric Tonnelier, Nicolas Baskiotis, Vincent Guigue, and Patrick Gallinari

Anomaly Detection in Smart Card Logs and Distant Evaluation with Twitter: a Robust Framework

In this paper, the authors present four approaches for the task of anomaly detection in a transportation network using smart card logs. In particular, data coming from the Parisian metro network, composed of 300 stations and millions of daily trips, is considered as a case study. In order to evaluate the proposal, authors ana-

lyzed particular days, but they also mined the Parisian transport authority (RATP) Twitter account to obtain (partial) ground truth information about operating incidents. Results show that matrix factorization, one of the four proposed approaches, is very robust in various situations while the last user-based model, another method proposed by the authors, is particularly efficient to detect small incidents reported in the twitter dataset.

- iii) Benjamin Paassen, Alexander Schulz, Janne Hahne, and Barbara Hammer

Expectation Maximization Transfer Learning and its Application for Bionic Hand Prostheses

In this paper, the authors propose a novel approach to handle changes in the distribution of data coming from the domain of bionic hand prostheses, where machine learning models promise faster and more intuitive user interfaces, but are hindered by their lack of robustness to everyday disturbances, such as electrode shifts. In order to address changes in the data distribution, authors exploit transfer learning, that is, to transfer the disturbed data to a space where the original model is applicable again. In particular, a novel expectation maximization algorithm to learn linear transformations that maximize the likelihood of disturbed data according to the undisturbed model is proposed. The authors also show that this approach generalizes to discriminative models, in particular learning vector quantization models. Results demonstrate that the proposed approach can learn a transformation which significantly improves classification accuracy and outperforms current baselines.

- iv) Olli-Pekka Rinta-Koski; Simo Särkkä, Prof.; Jaakko Hollmén, Markus Leskinen, Sture Andersson

Gaussian Process Classification for Prediction of In-Hospital Mortality among Preterm Infants

This paper presents a method for predicting preterm infant in-hospital mortality using Bayesian Gaussian process classification. The authors combine features extracted from sensor measurements, made during the first 72 hours of care for 598 Very Low Birth Weight infants of birth weight $< 1500\text{g}$, with standard clinical features calculated on arrival at the Neonatal Intensive Care Unit. Time periods of 12, 18, 24, 36, 48, and 72 hours were evaluated. The proposed method achieve a classification result with area under the receiver operating characteristic curve of 0.948, which is in excess of the results achieved by using the clinical standard SNAP-II and SNAPPE-II scores.

The guest editors would like to thank all authors for their interesting contributions and all reviewers for their excellent work. Authors and re-

viewers were asked to respect a very tight schedule, which allowed this issue to be published in less than a year after the conference, timely before the ESANN meeting of 2018. We would also like to thank the Neurocomputing editorial board for giving us the opportunity to publish this issue, as well as Elsevier's people for the very efficient and seamless management of the publication procedure. Finally, our most sincere gratitude goes to Prof. Michel Verleysen for his strong support of this special issue and his excellent conference organization.

Guest Editors

Fabio Aiolli

Dipartimento di Matematica, Università degli Studi di Padova, Italy

Michael Biehl

Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands

Luca Oneto*

Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi, Università degli Studi di Genova, Italy

*Corresponding author.

Fabio Aiolli received a PhD in Computer Science in 2004 from the University of Pisa. He was Post-doc at the Computer Science Department, University of Pisa (Italy), Visiting Scholar at the University of Illinois at Urbana-Champaign (IL), USA, and Post-doc at the Dept. of Pure and Applied Mathematics, University of Padova (Italy). Since 2006, he is Assistant Professor at the Dept. of Mathematics, University of Padova (Italy). His research activity is in the area of Machine Learning and Pattern Recognition. In particular, his expertise is in kernel methods for structured data, kernel and representation learning, hierarchical representations and deep learning, with applications to recommender systems, neuroscience and biology.

Michael Biehl received a PhD degree in Physics from the University of Gießen, Germany, in 1992 and the habilitation (*venia legendi*) in Theoretical Physics from the University of Würzburg, Germany, in 1996. He joined the University of Groningen as Assistant Professor in Computer Science in 2003 and got tenure in 2009. His earlier research concerned, among other topics, the statistical physics of neural networks and the theory and simulation of non-equilibrium physical systems. More recently, his interests are focused on the development and study of advanced machine learning methods and their practical application, for instance in the bio-medical domain.

Luca Oneto was born in Rapallo, Italy in 1986. He received his BSc and

MSc in Electronic Engineering at the University of Genoa, Italy respectively in 2008 and 2010. In 2014 he received his PhD from the same university in School of Sciences and Technologies for Knowledge and Information Retrieval with the thesis "Learning Based On Empirical Data". In 2017 he obtained the Italian National Scientific Qualification for the role of Associate Professor in Computer Engineering. He is currently an Assistant Professor at University of Genoa with particular interests in Statistical Learning Theory, Machine Learning, and Data Mining.